IRANIAN EFL LEARNERS' ATTITUDE, PREFERENCE, AND FAMILIARITY WITH TEST PERFORMANCE ON CBT AND PPT

MONIROSADAT HOSSEINI

UNIVERSITI SAINS MALAYSIA
2015

IRANIAN EFL LEARNERS' ATTITUDE, PREFERENCE, AND FAMILIARITY WITH TEST PERFORMANCE ON CBT AND PPT

By

MONIROSADAT HOSSEINI

Thesis Submitted in Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

February 2015

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and appreciation to the following people, without their help and support, this venture would not have reached completion. But first and foremost, I would like to express my deepest thanks to Allah, the Almighty for giving me wisdom, opportunity, and strength to start and complete this endeavor. His strong will toward successful accomplishment of higher degrees has been and is eternally bestowed upon my life

I am heartily indebted to my supervisor, Dr. Mohamad Jafre Bin Zainol Abidin, whose amicable guidance and support from the initial to the final stage of the study encouraged me to accomplish this project. He was always ready to help me and paid meticulous attention to details of the research in spite of his tight schedule as the Program Chairman for TESOL at the School of Educational Studies, University Sains Malaysia. I could not have completed this work without his support and supervision.

I would also like to extend my special thanks to my co-supervisor, Dr, Manoochehr Jafari Gohar, whose guidance, generosity, encouragement, and advices were a source of encouragement and motivation for me as I was completing my studies in both Iran and Malaysia. His guidance and supervision has started since my graduate studies ten years ago.

I owe a special note of gratitude to my co-supervisor, Dr. Amelia Bint Abdullah for the assistance and advice I received from her throughout the study. Her patient attention to my work inspired me to do my best to complete this work more carefully.

Furthermore, I need to express my Heartfelt thanks to Professor Dr. Abdul Rashid Mohamed, Dean of the School of Educational Studies, Associate Professor Dr. Abdul Rashid Mohamed, and Dr. Shaik Abdul Malik Mohamed Ismail, Deputy Deans, whose advice during the study guided and inspired me in going ahead in this journey. I also express my special gratitude to the Doctoral Committee Members for their patience in reading and giving invaluable comments for improving this work. I would also like to express my thanks to the staffs of the school for their efforts and help in completing the work on due time, especially Mrs. Faridah Bharum and Nur Hidayah Abdul Halim.

I am also highly grateful to Dr. Alizadeh, Dr. Oraki and Dr. Farahmand in Tehran Payame Noor Center for their kind cooperation in collecting data and granting me permission to access the participants, venues, and required facilities. Surely, without their help, I could not collect required data. Moreover, I am extremely thankful to all the participants who provided me rich and detailed data for the study and lent breadth and value to the research findings.

My special thanks are also to all my colleagues at Payame Noor University, Tehran, Iran particularly to Dr. Salim Bahrami, Mrs Hosseini Raoof, Mrs Eshaghi, Mr, Yaseri Far, Mrs. Neda Shahin Pour, and Ms. Mohtashami for their priceless cooperation and commitment in assisting me through the process of collecting data. I would like to express my thanks to my colleagues in International Center of Payame Noor University who supported me in my work, Mr. Masood Piri, Farah Rahmati Moghaddam, Fattane Vali Nejad, Maryam Ziaeeian, Fereshteh Falahat Doost, and Monire Beheshti.

My sincere thanks also go to Dr. Abolfazl Shirban, Dr. Majid Poor Mohammadi, Dr. Vahid Rafieyan, Dr, Hamid Reza Keshavarzian, Dr Mohammad Mahmoudi Meimand, and Dr. Hamid Kamarzarin whose inspiring and helping in so many ways encouraged me in the way of this venture.

I am also deeply thankful to my parents and siblings who deserve special mention for their support, prayers, and perpetual presence in my life, visible or invisible. I know they would be very proud of my success in this endeavor.

Last but not least, I wish to express my deepest gratitude wholeheartedly to my unfailingly kind husband, Ahmad Asanjarani Farahani and my dear sons, Behrooz and Behzad who have endured me and whose prayers, love, and best wishes helped and encouraged me throughout the period of my study. Actually words fail to express my deepest thanks to them as without their patience, understanding, and standing behind me, I could never pursue and complete my study. Thank you for standing beside and behind me.

Dear Ahmad, Behrooz, and Behzad, Thanks millions for all your supports.

Monirosadat Hosseini

February 2015

TABLE OF CONTENTS

CHAPTER II. LITERATURE REVIEW

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASVAB	Armed Services Vocational Aptitude Battery
CALL	Computer Assisted in Language Learning
CALT	Computer Assisted in Language Testing
CAT	Computer adaptive Test
CBT	Computer Based Test
C-TAM-TPB	Combined TAM and TPB
DAT	Differential Aptitude Test
EFL	English as a Foreign Language
ESL	English as a Second Language
GRE	Graduate Research Examination
IDT	the Innovation Diffusion Theory
ITED	Iowa Tests of Educational Development
MAB	Multidimensional Aptitude Battery
MCQs	multiple-choice questions
MM	Motivational Model
MPCU	Model of PC Utilization
NAEP	National Assessment of Education Progress
OST	Office Skills Test
PLS	Partial Least Squares
PNU	Payame Noor University
PPT	Paper and Pencil Test
SAD	System of Administering and Developing tests
SCT	Social Cognitive Theory

SPSS	Statistical Package for Social Science
TAM	Technology Acceptance Model
TPB	Theory of Planned Behavior
TRA	Theory of Reasoned Action
TOEFL	Test of English as a Foreign Language
UTAUT	Unified Theory of Acceptance and Use of Technology
WPT	Western Personnel Test

LIST OF PUBLICATIONS DURING THE PHD EDUCATION

Abidin, M. J. Z., Osman, S. R. F., Hosseini, M., (2012). Role-Play: Taking the Line of Least Resistance. *International Journal of Learning & Development*. 2(2), 258-270.

Hosseini, M., Abidin, M. J. Z., Kamarzarrin, H., & Khaledian, M. (2013). The investigation of Difference between PPT and CBT Results of EFL Learners in Iran: Computer Familiarity and Test Performance in CBT. *International Letters of Social and Humanistic Sciences*, 11, 66-75. Available online at www.ilshs.pl.

Hosseini, M., Abidin, M. J. Z., Taghizadeh, M . T., &Naseri, E. (2013). In the Importance of EFL Learners' writing Skill: Is there any Relation between Writing Skill and Content Score of English Essay Test?*International Letters of Social and Humanistic Sciences*, 6, 1-12. Available online at www.ilshs.pl.

Hosseini, M., Abidin, M. J. Z., Baghdarnia, M. (2014). Comparability of Test Results of Computer Based Tests (CBT) and Paper and Pencil Tests (PPT) among English LanguageLearners in Iran. *Procedia - Social and Behavioral Sciences* 98 659 – 667. Available online at www.sciencedirect.com, Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Abidin, M. J. Z., Hosseini, M., Choobdari, B. (2014). Iranian Students' Perception on Intensive English Courses of USM: Is it Helpful in Getting Good Score in IELTS? Malaysian International Conference on Languages, Literatures & Cultures, MICOLAC (2014), Malaysia, Penang. University Putra Malaysia (UPM), 63.

KEUTAMAAN, KEBIASAAN DAN SIKAP PELAJAR BAHASA INGGERIS IRAN TERHADAP DALAM UJIAN PENCAPAIAN KOMPUTER (CBT) DAN UJIAN PENCAPAIAN PENSIL DAN KETAS (PPT)

ABSTRAK

Kajian ini bertujuan untuk mengkaji kesataraan hasil Ujian Pencapaian Komputer (CBT) dan Ujian Pencapaian Pensil dan Kertas (PPT) dengan membandingkan purata skor dua ujian mod terhadap pelajarbahasa Inggeris Pendidikan Jarak Jauh di University Payame Noor di Iran. Kajian juga ingin menyiasat hubungan antara kemungkinan sikap mereka terhadap penggunaan Komputer dalam ujian prestasi CBT dan korelasi di antara kebiasaan penggunaaan komputer dan hasil ujian terhadap CBT. Keutamaan mod ujian diambil kira dan diukur dalam kajian ini. Kajian menggunakan kaedah gabungan kuantitatif dan kualitatif melibatkan 202 orang pelajar Bahasa Inggeris pendidikan jarak jauh dari Tehran Payame Noor University Centre. Kesemua pelajar diberi dua ujian yang setara daripada Buku kursus Bahasa Inggeris, satu dalam mod pensil dan kertas, dan yang selebihnya melalui CBT. Setiap ujian mod mempunyai selang masa empat minggu. Soal Selidik Skala Likert yang mengandungi dua bahagian yang digunakan sebelum kajian dilakukan diberi kepada pelajar. Bahagian pertama daripada soal selidik diadaptasi dari Computer Attitude Scale (CAS) dihasilkan oleh Loyd dan Gressard (1985) digunakan untuk mendapat maklumat mengenai tingkah laku pelajar. Temuduga terhadap kumpulan fokus seramai 14 orang pelajar digunakan untuk melihat sebab mengapa wujud perbezaan di antara PPT dan CBT. Analisis kuantitatif menunjukkan perbezaan yang signifikan terhadap skor purata PPT dan CBT di mana skor lebih tinggi terhadap PPT. Walau bagaimana pun tidak terdapat perbezaan yang signifikan dalam skor CBT terhadap jantina dan umur. Hasil kajian juga

menunjukkan tiada hubungan yang signifikan terhadap tingkah laku pelajar yang menggunakan komputer dan kebiasaan mereka terhadap komputer pada CBT. Tiada juga hubungan yang signifikan di antara tingkah laku pelajar yang lebih berusia dan pelajar yang lebih muda. pelajar lelaki menunjukkan tingkah laku yang lebih positif dan pelajar yang berusia muda menunjukkan sikap kebiasaan yang lebih terhadap penggunaaan komputer. Analisis kualitatif menunjukkan sebab responden mengutamakan penggunaan CBT. Maklumat tambahan menunjukkan mereka lebih gemar dengan cirri-ciri CBT tetapi baik dalam penggunaan PPT. Secara keseluruhan, hasil kajian menunjukkan era dalam kemajuan penggunaan teknologi di mana penggunaan komputer diperlukan dalam pendidikan peringkat rendah dan pembelajaran untuk memastikan penggunaan yang dioptimakan. Hasil kajian juga menunjukkan pembelajaran dan instrumen pembelajaran perlu digunakan bergandingan bagi mendapatkan hasil yang lebih diinginkan. Dari kajian ini juga menunjukan bahawa perubahan pentadbiran mod ujian dari tradisi kepada yang lebih moden dan berteknologi melalui penggunaan komputer dalam sistem pendidikan jarak jauh harus dikaji dari masa ke semasa untuk mengurangkan jurang perbezaan di antara CBT dan PPT. Pengalihan CBT kepada PPT tidak dapat dilelakkan

IRANIAN EFL LEARNERS' ATTITUDE, PREFERENCE, AND FAMILIARITY WITH TEST PERFORMANCE ON CBT AND PPT

ABSTRACT

This study investigates the comparability of test results in Computer Based Test (CBT) and Paper-and-Pencil Test (PPT) by comparing the mean scores of two test modes among distance English language learners of Payame Noor University (PNU) in Iran. Besides, it seeks to probe the probable relationship between their attitude towards the use of computer and test performance on CBT. In addition, it attempts to examine the correlation between participants' computer familiarity and their test results on CBT. It also seeks to discover the test mode preference of test takers and their results on CBT. This research employs a mixed method design. A total of 202 English language learners from Tehran Payame Noor University Center participated in this study. All participants were administered two equivalent tests from the General English course book, one in paper-and-pencil mode, and the other in computer-based mode with the interval of four weeks. After conducting two equivalent tests, a Likert rating-scale questionnaire consisted of two parts, which was piloted before main study, was given to all participants. The first part of the questionnaire, adapted from Computer Attitude Scale (CAS) developed by Loyd and Gressard (1985), was used to elicit information about attitude of participants. Focus group interview was conducted among 14 students to discover the reasons for their test mode preference for PPT or CBT. The results of quantitative analysis showed significant difference between mean scores of two tests with the higher score on PPT. However, it was shown that there is no significant difference in CBT scores between two gender and age groups. Besides, the research revealed no significant

relationship between attitude towards the use of computer and computer familiarity with test results on CBT. It also found no significant difference between the level of attitude among older and younger groups; nonetheless, the male group showed higher positive attitude, and younger students showed higher familiarity with computer. In qualitative analysis, respondents justified the reasons of their preference for CBT. In addition, they revealed generally more preference for CBT but did better on PPT. Overall, the results supports that in the era of technology development, it is necessary to include computers in lower level of educations and learning contexts to make using it more ordinary. The results implied that teaching and testing instruments should be in parallel with each other to get a desirable result from both teaching and testing. The findings indicated that transition from traditional test administration mode to more modern and technological mode through computers in distance educational systems needs more investigations and time to cope with the barriers for equivalency of two tests. Moreover, the substitution of CBT for PPT is inevitable in the future, especially in language teaching contexts.

CHAPTER I

INTRODUCTION

1.1 Overview

The necessity of using technological devices both in learning and testing in educational settings have been rapidly increased since widespread accessibility to computers and broad developments in information and communication technology (Bachman, 2000; Chapelle, 2007). The widespread use of computer technology in conducting language tests and the availability of personal computers, along with increased computer familiarity are making computer-based test administration feasible on a large scale. In this regards, computer based testing is going to be applied all around the world in academic contexts (Chappel & Douglas, 2006; Clariana & Walace, 2002). Developments in language testing studies during the past years provided evidence in making use of technological devices in language testing mostly in test developing, administering, scoring, reporting and processing given data (Douglas, 2000).

Meanwhile, increased accessibility to personal computers, widespread use of computer in language learning, increased computer familiarity, and willingness to use computers in educational settings motivated researchers to conduct studies considering these factors in comparability studies (Fazeli, Ross, Vace, and Ball, 2013; Lightstone & Smith, 2009; Maguire, Smith, Brallier, & Palm, 2010; Salimi, Rashidy, Salimi, & Amini Farsani, 2011; Terzis & Economides, 2011; Yurdabakan & Uzunkavak, 2012).

Given the integral role computers play in our lives, the number of computer-delivered tests is increasing in language testing due to the perceived advantages of computer-delivered tests (Lottridge, Nicewander, Schulz, & Mitzel, 2008; Paek, 2005). Such developments in technologies have influenced many areas in educational settings such as online learning and testing (Bennett, 2002; Dooling, 2000; Pommerich, 2004).

In addition, in language learning, the use of computers and electronic devices has become popular; especially in assessing the language proficiency of English learners, the most precise and available way could be through computers (Bachan, 2000; Clariana & Wallace, 2002; Paek, 2005; Sawaki, 2001). However, the high cost of using computer in high stake tests and less computer familiarity had limited the implementation of computerized language testing (Clariana & Wallace, 2002).

On the other hand, as computers become increasingly available in distance educational settings, it is likely that examiners use them to administer tests (Trotter, 2001). In the new decade, the influence of computers and technology on distance educational settings is emerging as an efficient tool for delivering instructional content and is drastically expanding in the area of student assessment (Ibid). McFarland (2001) proposed that computer-based tests provide advantages over traditional paper-and-pencil tests even if the computer version is a simple, non-adaptive replication of the paper version.

Bugbee and Brent (1990) asserted that once set up, CBT is easier to administer than (Paper-and-Pencil Test) PPT and offer the possibility of instant grading. Inouye and Bunderson (1986) suggested that in CBT, testing conditions can be standardized and the sequence of items can be easily manipulated. These advantages and the like made administrators in distanbce educational settings include (Computer-Based Test) CBT in their system gradually to employ technology in their systems (Trotter, 2001).

Despite the enthusiasm and advantages of administering CBT, research has shown many barriers to the use of computers in education including the lack of training and support, instructor self-efficacy, and computer anxiety (Clarian & Wallace, 2002; Trotter, 2001).

1.2 Background

One of the most appropriate ways of measuring students' learning in educational setting is assessment (Bachman, 2000). Portfolio assessment, performance assessment, self-assessment, and peer assessments are among the examples of different types of assessment (Peat & Franklin, 2002). In recent years, information and communication technology has been employed in assessment and examination to mechanize the testing process. Computer-Based Testing (CBT) provides a variety of innovations in testing and can be used in different contexts; one of the important areas is language testing (Bennet, 1998). The history of computerized testing began in the early 1970s (Bachman, 2000; Bunderson, Inouye, & Olsen, 1989; Chapelle, 2007; Mazzeo & Harvey, 1988; Mead & Drasgow, 1993; Wainer, Doran, Flaugher,

Green, Mislevy, Steinberg, & Thissen, 1990). With the appearance of new technologies, computerized testing has begun to be widespread and implemented in large-scale tests (Higgens, Russell, & Hoffmann, 2005). Examples include state drivers' license exams, military training exams, job application exams, entrance exams in postsecondary education, and certification exams by professional groups such as TOEFL or IELTS (Russo, 2002; Trotter, 2001).

The limited accessibility to computer and high cost limited the implementation of computerized language testing in past years (Anakwe, 2008; Chapelle & Douglas, 2006; Paek, 2005); however, recent developments in communication technologies have created alternative test methods through computers and internet all around the world (Clariana & Wallace, 2002).

Since 1990s, many researchers advocated the importance of assessment in helping students learn better (Earl, 2003; Hart, 1994; Leahy, Lyon, Thompson, & William 2005; Marzano, Pickering, & McTighe, 1993; Persichitte & Herring, 2002; Popham, 2002; Wiggins, 1993). Earl (2003) describes examination in today’s schools as primarily evaluation of learning. On the other hand, computerized testing advocates believed that traditional measurement implementation place too much emphasis on passing a test rather than on encouraging learners to learn beyond education (Tanner, 2001).

However, as institutions started to accomplish computer-based testing in their examination systems, concerns came up about the comparability of scores from the two administration modes, PPT and CBT (Chapelle & Douglas, 2006; Clariana & Wallace, 2002; Wang, 2004). As the computerized tests have been using for almost 20 years (Laborda, 2007), and the Computer Assisted Language Learning (CALL) has been common since the middle of 20th century, it has been necessary to develop the means to include computerized tests in language testing (Leahy et al., 2005).

1.2.1 Computers and Language Testing

Although computer has played an important role in testing for more than 20 years, the literature on CALL has shown that there has been relatively little attention to Computer Assisted in Language Testing (CALT) (Bachman, 2000; Sawaki, 2001). While computers have been important in language assessments, only a relatively small group of professional language testers used computers in producing and validating language tests (Sawaki, 2001).

Russell and Haney (2000) asserted that the ″mismatch between the mode of learning and assessment could cause achievement to be inaccurately estimated.″ (p.2).

As computers become increasingly available in educational settings, it is likely that English teachers will use them to administer tests (Trotter, 2001). Bennet (2002) believes that since computers entered in our lives and

had integral role in education, and as developments in technology made measurement of constructs more possible, it is clear that the use of CBT for language testing will become increasingly inevitable in the future (Bennet, 2002). However, Norris (2000) raised the question about the comparability of CBT and PPT in language testing in that whether CBT can provide appropriate means to interpret the language skills or proficiencies tested according to language educators' purposes, and whether it fulfills the intention of language testing uses.

Although CBT offers many advantages over traditional PPT (Lottridge et al., 2008; Poggio, Glasnapp, Yang, & Poggio, 2005; Russel & Haney, 1996; Sawaki, 2001; Zhang & Lau, 2006), assessment experts, researchers, practitioners, and educators have concerns about the equivalency of scores between the two test administration modes (Chapelle & Douglas, 2007; Douglas, 2000; Lottridge et al., 2008). To deal with this concern, many researchers conducted studies in synthesizing the administration mode effects on CBTs and PPTs (Clariana & Wallace, 2002; Higgins et al., 2005; Johnson & Green, 2006; Olsen, Maynes, Slawson, & Ho, 1989; Paek, 2005; Poggio et al., 2005; Pommerich, 2004; Salimi et al. 2011; Zandvliet, 1997; Zhang & Lau, 2006). Some researchers found that in comparability studies on CBT and PPT, test takers have done better on CBT (Bugbee & Bert, 1990; Clariana & Wallace 2002; Lee, Osborne, & Carpenter, 2010; Maguire et al., 2010; Parshall & Kromery, 1993) and in some others test takers performed better on PPT (Al-Amri, 2008; Anakwe, 2008; Pomplun et al, 2002; Salimi et al., 2011).

1.2.2 Comparability Studies in Language Tesing

Due to the variety of results of similar studies and the necessity of substituting CBT for PPT in some educational settings, especially in distance educational systems, where using electronic devices is inevitable, conducting comparability study is vital (Lottridge et al., 2008). Comparability of test scores should be examined before replacing or including CBT in the language assessment procedures (Pommerich, 2004). In addition, as computerized testing could be affected by the students' intention behavior in using computer and their attitude and preference, doing some studies considering these variables in test results is important to see whether various testing modes examine the same construct without the interference of other irrelevant variables. For example, the literature on computer-assisted language learning (CALL) indicates that both language learners and instructors have generally positive attitudes toward using computers in the classroom, but the evidence of their attitude towards the specific area of computer use, i.e. language testing is not enough (Esmaiel, 2006; Stricker & Attali, 2010). One of the barriers in implementing computer based tests among English language teachers in the university could be related to their students' reluctance in substituting new instruments in examination (Yurdabakan & Uzunkavan, 2012). Reece and Gable (1982) found that identification of students with positive or negative attitudes towards the use of computers as well as methods that may influence those attitudes should be of great value for curriculum planners and program evaluators.

In recent years, computer-based assessments have grown in popularity internationally and are increasingly being used in developed countries. Due to their accuracy and speed of execution, they are predicted to become the primary mode of assessment in the future in developing countries (Wang, et al., 2007).

There have been studies on comparability of test results in PPT and CBT considering key factors associated with test results in different countries with different languages and technological backgrounds (Al-Amri, 2009; Bachman, 2000, Busch, 1995; Flowers et al., 2011; Chappel, 2007; Douglas, 2000; Yurdabakan & Uzunkavan, 2012). Nevertheless, there are disparities in the results of such studies. Thus, studies are necessary to investigate the factors affecting the students' acceptance and intention to use CBT as well as their actual behavior in it.

Fulcher (1999) suggested that not only the issue of equivalency is important in such comparability studies, but also other equating issues such as previous familiarity with using computers, attitudes towards the use of computers and CBT, and their backgrounds are crucial to consider. Some studies showed that students have high interest in using computers, but are reluctant in doing their exams by computer.

1.2.3 Advantages of Computer-based Tests

Noyes and Garland (2008) believe that the benefits of standardized computer-based testing, such as quick and objective results as well as the ease of reporting results to others make this method very popular. Moreover, moves towards computerized testing is rooted from the advantages it provides in comparison with traditional paper-and-pencil format (Neuman, & Baydoun, 1998; Pomplun & Custer, 2005; Salimi et al., 2011; Terzis & Economids, 2011; Yurdabakan & Uzunkavan, 2012). Such advantages, according to the findings of mentioned studies, include cost-effective administration, ease of administration, more accuracy, immediacy of scoring and reporting, and flexible test scheduling. These studies, also, indicated that students who are familiar with computer feel more comfortable while using it (DeBell & Chapman, 2003; Higgens et al. 2005; O'Malley, Kirkpatrick, Sherwood, Burdick, Hsieh, & Sanford, 2005; Poggio et al., 2005).

Because of its advantages, computerized testing now plays an important role in educational assessments (Clarian & Wallace, 2002; Kingston, 2009; Poggio et al., 2005; Russel & Haney, 1996). It should be noted that some disadvantages are attributed to computerized testing as well, namely the higher costs of item development, which can outweigh many of the savings gained through the advantages (Noyes & Garland, 2008). A well-designed computer program can display the assessment tasks to candidates, mark the responses, grade responses, record and instantly provide feedback on the performance of each candidate. Moreover, a series of password codes can

completely bar unauthorized entry to the assessment tasks to those with intents to temper with candidates' scores.

The careful look at the discussions of some comparability studies showed that they emphasized on their suggestions on the need of more careful investigation on the even weak differences in scores between CBT and PPT. They believed that some construct-irrelevant variables could influence the results of computerized test (Al-Amri, 2009; Bachman, 2000; Busch, 1995; Chappel, 2007; Douglas, 2000; Stricker, Wilder, & Rock, 2004) and should be considered important in similar studies.

1.3 Rationale of the Study

Using computers in language testing as well as in language learning has some advantages and disadvantages. The interrelationship between computers and test takers' characteristics, such as attitude towards the use of computer, familiarity with computer, and test mode preference could influence the comparability and interpretation of the scores of both PPT and CBT (Al-Amri, 2008). However, the issue of equivalency of PPT and CBT has been arisen due to certain differences in administration as well as differences in test taker characteristics (Clariana & Wallace, 2002; McDonald, 2002; Sawaki, 2001). Despite the widespread use of computer-based testing, relatively few studies have been conducted on the equivalency of two test modes in distance educational setting. However, some institutes and educational settings are going towards using computerized test due to its advantages without doing any comparability investigation beforehand.

Perhaps because they mostly believe that if the items are identical, the testing mode is irrelevant (Linghstone & Smith, 2009).

In addition, CBT can be used in different contexts because it provides a great deal of prospects for innovations in testing and assessments (Bennett, 1998). In recent years, CBT has become very popular for providing advantages to the academics and practitioners such as test security, cost and time reduction, speed of reporting the results, automatic processes in recording and keeping for item analysis, and especially for distance learning accessibility to test procedures (Bugbee, 1996; Smith & Caputi, 2005).

Payame Noor University (PNU), as the only distance educational system in Iran, needs to implement technological devices in educational settings and computer in language teaching and testing. Entering computers in educational contexts in PNU, staffs, teachers and students are encouraged gradually to start using the new system in their teaching, learning, and assessment tasks. These procedures were introduced without examining the readiness of the target context, the pedagogical effect on students, or the consequences on the students' achievement.

In addition, there are some research comparing CBT with PPT in ordinary educational settings (e.g., Pomplun, Frey, & Becker, 2002; Wang et al., 2008). However, very few studies have been done for the comparability

of CBT and PPT in language testing in distance educational systems, in this case PNU in Iran (Barootchi and Keshavarz).

On the other hand, one of the goals of using computerized testing should be providing students the ability to apply the available technology devices adequately, while not all students have the same confidence in being successful in using technology appropriately in educational settings especially in examination (Al-Amri, 2009).

In addition, conducting computer-based tests could be useful in reducing students' anxiety through examination and could give them the sense of progress in the era of technology development. Moreover, familiarity with computerized test format could prepare such distance learners for taking high stake tests such as TOEFL, IELTS, GRE or the like. Moreover, testing and teaching materials should be in parallel with each other to get a desirable result from both procedures because according to Russell and Haney (2000) mismatch between the mode of learning and assessment could cause achievement to be inaccurately estimated.

Besides, knowing how to prepare for and take exams can affect students' attitudes towards exam, reduce their test anxiety, build up their self-confidence in exam scores, and ultimately, help them to obtain better grades (Holzer, Madaus, Bray, & Kehle, 2009). The users' attitudes, familiarity, and preference may play an important role while taking test on computers. Some

studies on the comparability of CBT and PPT have been conducted considering such test mode factors whose results are inconsistent to be referred to by the instructors of the university. Some studies showed that the differences between test scores of two test modes are mostly due to the students' attitude towards the use of computers during the testing process (Bachman, 2000, Busch, 1995; Douglas, 2000; Yurdabakan & Uzunkavan, 2012).

In summary, the reason of conducting this study is the necessity of investigating the equivalency of both exams considering some important examinees' characteristics affecting test results, such as attitude towards the use of computer, familiarity with computer, and test mode preference based on Technology Acceptance Model (TAM) considering the importance of substituting CBT for PPT in PNU. It is worth mentioning that as the criteria of students' promoting to higher educational levels in PNU is final exam results, this type of investigation could be applicable for educators and decision makers. Therefore, a practical comparison between these two testing modes needs to be carried out to determine whether the testing mode has considerable influence on the test results of PNU English language learners. The reason why this university has been chosen is that it is the biggest and the only university under the name of distance education system in Iran. Under such system, challenges have recently emerged whereby examination by papers have leaked before the examination date, administration costs have skyrocketed, and the manual assessment of the multiple-choice questions (MCQs) have been blamed for causing untold delays in the publication of the

results. In line with modern developments in science and technology, computer programs have the potential to alleviate some of these challenges at a minimum cost.

In this university, because of the high rate of students all over the country and other countries, administering final tests is very time consuming and costly for both students and administrators. Moreover, PNU is a distance education system and the results of tests could be practical because the only way of evaluating students is final examinations, thus considerable attention should be given to improve the test administration conditions appropriate to distance educational systems. It means, if students do not experience English exams through computers in the local academic contexts, they cannot perform easily when encounter with standard global English language on-line exams such as TOEFL, IELTS, GRE, etc.

1.4 Statement of the Problem

Distance education institutions need to go through a gradual process of transition from traditional testing to more technologically-based testing with the result of saving considerable amounts of time, energy and money on the part of both students and institutions. In Payame Noor University, all the summative examinations, set by the central examination department of the university, are conducted in paper-based and recently computer-based testing types. Administering computer-based exam in Payame Noor University since 2011 through system of administering and developing tests (SAD) system, yielded questions about the interchangeability of the results of two test modes

while there has not been any investigation considering this issue. Moreover, the researchers in the field have controversies whether test takers perform the same on CBT and PPT. It is necessary to refer to the stable results derived from comparability studies on transiting PPT to CBT. However, to the best of the researchers' knowledge, there is not any comparability study before including or substituting CBT for PPT in PNU. Dillon's (1992: iv) believes that "the assumption of comparability between CBT and PPT without proper investigation within that particular testing context, is inappropriate".

On the other hand, the anecdotal evidence collected through interacting with PNU students seems to suggest that many of them do not like fully automated tests; neither do they place sufficient trust in the validity of such tests. The investigations of test takers' attitudes to CBT can therefore help us clarify whether such anecdotal evidence holds true. They claimed that when they face with computerized testing, they feel anxious about the process. It has been argued that the mode of test administration somehow could affect the scores of test takers because in the case of any difference in test scores of two modes, it can most likely be attributed to the willingness of the students in using computer or doing electronic exams that should be considered in comparability studies (Parshall & Kromrey, 1993). Some students might have generally positive attitudes toward using computers in the classroom, but the evidence of their attitude towards language testing is not enough (Esmaiel, 2006; Stricker & Attali, 2010).

Based on TAM theory (Davis, 1989), not only having positive attitude towards the use of computer could influence the performance on English language computerized tests, but also the good performance on tests also, by itself, could affect test taker's attitude towards the use of computers which in turn could affect the actual use of technology in language learning and testing**.** Murray et al. (2012) argue that a test taker's attitude towards a language test forms part of test impact, which is one of the essential qualities in Bachman and Palmer's (1996) test usefulness framework and described by Shohamy (1998) as consequential validity. Messick (1989) also explicitly recommends that attitude be considered as a crucial source of evidence for construct validity. Therefore, as test takers' attitude is believed to be part of test impact, a fair test should function equally among different groups of test takers.

In addition, some students after conducting CBT in PNU complain about their lower test scores due to unfamiliarity with the new test administration mode and claim that their grade exams are not what they could get on more familiar PPT format. Some also believe that the familiarity with computer could be a major factor influencing their test performance on CBT. The results of the effect of computer familiarity on test performance are inconclusive; some studies showed different results in finding out the relationship between variables of computer familiarity and test results in CBT (Al-Amri, 2009; Odo, 2012). Some researchers, in their conclusions of their studies, claimed that the reason of this inconsistency could be due to the gender and age differences in relation to attitude and computer familiarity

(Al-Amri, 2009; Bachman, 2000; Busch, 1995; Yurdabakan & Uzunkavak, 2012).

Some studies concluded that examinees' preference in choosing the test delivery format can be effective in test performance (Al-Amri, 2009; Lightstone & Smith, 2009; Mojarrad et al., 2012). Test administrators and decision makers of the university should consider this issue because it may distort or invalidate the test takers' performance when the only criterion for students' progress to upper level in this university is the results of final exams. Therefore, it is necessary to do an investigation considering such related variables to test administration modes in PNU to investigate the effect of test takers' preference and their oerformance on both tests of CBT and PPT.

As female students in the target context represents about 75% of the overall students, the issue of difference in performance between males and females should be considered carefully. In addition, female students complain more in doing CBT than male students do and show less confidence in doing it. They also tend to show less familiarity with computer to avoid such automated tests. So investigating the scores of famales and males in terms of attitude, familiarity, and CBT scores and their difference is inevitable in this context.

While computer using has increased over the past decades, studies have shown that older students in general show poorer computer attitude, less computer familiarity, and higher computer anxiety than the younger (Laguna

& Babcock, 1997; Fazeli et al., 2013). Such findings imply the influence of age differences on attitude as well as on computer familiarity and test results in CBT. As the age range of students in PNU is between 18 and 65 due to its context, the results of this study might be influenced by the age of participants while examining the relationship between variables of familiarity and attitudes towards the use of computer and computerized test scores. In addition, as the slogan of the University is 'education for anybody, anytime, anywhere', the issue of age and gender differences is a crucial factor to be considered by decision makers and university stakeholders to provide final test as fair as possible.

Most importantly, in order to get a reasonable feedback from language testing, it is important that test takers concentrate more on the content of test not the way of doing tests. Poggio et al. (2005) emphasized the clear need of systematic study and exclusive investigation to make decision and direction at this time and states, "When a CBT system is implemented, it is paramount that examinees' responses are affected only by test content, not administration mode" (p. 5).

1.5 Research Objectives

The main objectives of the current tudy are:

1- To examine whether there is any significant difference in test results of PPT and CBT among EFL learners in PNU,

1.1 To test the difference in test results of CBT between older (36-above) and younger (18-35) EFL learners in PNU,

1.2 To test the difference in test results of CBT between female and male EFL learners in PNU,

2- To investigate the relationship between PNU English language learners' computer attitude and their test results on CBT,

2.1 To examine the difference between test results on CBT and attitude among older and younger students,

2.2 To examine the difference between test results on CBT and attitude among females and males,

3 To find out the relationship between PNU EFL learners' computer familiarity and their test results on CBT,

3.1 To examine the difference between test results on CBT and computer familiarity among older and younger students,

3.2 To examine the difference between test results on CBT and computer familiarity among males cand fremales, and

4 To investigate the relationship between test mode preference of PNU English language learners and their test performance on CBT and PPT.

In order to achieve these objectives, the research questions have been put forward.

1.6 Research Questions

To fulfill the objectives of the study, the following research questions have been arisen.

The main research questions are:

1- Is there any statistically significant difference in test results of computer-based and paper-based tests among EFL learners in PNU?

1.1 Is there any significant difference in test results of CBT between older (36-above) and younger (18-35) EFL learners in PNU?

1.2 Is there any significant difference in test results of CBT between female and male EFL learners in PNU?

2 Is there any significant relationship between PNU EFL learners' attitudes and their test results on CBT?

2.1 Is there any significant difference in test results on CBT and attitude between older and younger students?

2.2 Is there any significant difference in test results on CBT and attitude between males and females?

3 Is there any significant relationship between PNU EFL learners' computer familiarity and their test results on CBT?

3.1 Is there any significant difference in test results on CBT and computer familiarity between older and younger students?

3.2 Is there any significant difference in test results on CBT and computer familiarity between males and females?

4 What is the relationship between test mode preference of EFL learners in PNU and their test performance on CBT and PPT?

1.7 Research Hypotheses

Answering the above questions, this study will test the following null hypotheses:

H_0 1: There is no statistically significant difference in computer-based and paper-based test results among EFL learners in PNU.

H_0 1.1. There is no statistically significant difference in the result of CBT between older (36-above) and younger (18-35) EFL learners in PNU.

H_0 1.2. There is no statistically significant difference in the result of CBT between male and female EFL learners in PNU.

H_0 2: There is no statistically significant relationship between PNU EFL learners' attitudes and their test results on CBT.

H_0 2.1: There is no statistically significant difference in test results on CBT and attitude among older and younger students in PNU.

H_0 2.2: There is no statistically significant difference in test results on CBT and attitude among male and female EFL learners in PNU.

H_0 3: There is not statistically significant relationship between PNU EFL learners' computer familiarity and their test results on CBT.

H_0 3.1: There is no statistically significant difference in test results on CBT and computer familiarity among older and younger students.

H_0 3.2: There is no statistically significant difference in test results on CBT and familiarity among male and female EFL learners in PNU**.**

1.8 Significance of the Study

This study aims at contributing to a number of applied and theoretical domains of language testing. The results of this study could be important in different areas.

Firstly, the results could fill the gap of rare study on the comparability studies on PPT and CBT in the context of higher education in Iran, especially in distance educational systems where using technological devices is inevitable in the progress of the university.

Secondly, there is a lack of test comparability studies between PPT and CBT concerning L2 tests to identify any test mode effect (Chalhoub-Deville & Deville, 1999). The results of the study might clarify whether construct itself (language knowledge) or construct irrelevant variables (test mode factors) are relevant in test scores on CBT. If it is so, the results could be helpful for test developers and instructors to make better decisions to lessen such effects as much as possible.

Thirdly, from the results of previous studies, it is evident that most of comparability studies can be examined on two levels. First, comparability can be examined in terms of score equivalency. In other words, one can investigate whether the two modes - PPT and CBT- produce similar score distributions, such as similar means and standard deviations. Second, comparability can be examined in terms of construct equivalency (Lottridge et al., 2008) because construct comparability involves determining whether the tests in two modes are measuring the same construct to the same degree.

The results of this research will help English teachers, test developers, administrators, and other instructors to make better decisions about the testing methodologies they employ in testing English language. Even so computerized tests have no advantages over paper and pencil tests, the use of computer will help to save time, energy, and money for both institutes and learners and guarantee more security (Noyes & Garland, 2008). The results could provide supporting evidence in the necessity of substituting CBT for PPT in the context under investigation according to the point of view of students.

So this study is important because:

- It can provide the English language teachers and administers the necessary information for making better decisions in applying educational strategies with greater certainty regarding the using computer technology in English teaching and testing language.
- It informs university administrators, policy makers, supervisors, educators, and teachers using CBT advantages and disadvantages. This may be of value for the authorities to consider it and enhance the positive factors that influence the results of examination.
- The results of this study also could hopefully add to the literature and fill the research void in the area in general, and in PNU as distance educational system in particular, as a foundation for the research community to go ahead with further research on the curriculum developing, especially in English programs.

- With the current increase in computer-based testing, instructors and institutions, working distance-learning system, can be aware of and plan for possible test mode effects, especially in language teaching in distance educational systems such as PNU.
- The results can be useful to emphasize the importance of technology usage in the era of technology development. Technology usage might help language test developers move beyond conventional test design procedures that provide scores primarily to evaluate students. It also can facilitate designing a more systematic test approach, create interrelations among task characteristics and test taker's performances, and make inferences about underlying abilities and processes that are going to be measured.
- Distance education institutions need to go through a gradual process of transition from traditional testing to more technologically-based testing with the result of saving considerable amounts of time, energy and money on the part of both language students and institutions. The findings of this study can have pedagogical implications in decision making in this area.
- The results can also show the English teachers how the attitude of their students may affect their test results as well as their learning based on Technology Acceptance Model.
- The results may also help the instructors consider the familiarity with computers and taking tests by computer before including CBT on their curriculum.